

第 6 章：等式约束的优化问题

(1): 牛顿法

授课教师：曹语

课程主页：<https://yucaoyc.github.io/math3806>

为什么约束会改变牛顿法？

在无约束优化中，牛顿法直接朝局部二次模型的最小点走；但现在每一步还必须留在平面 $Ax = b$ 上。

本节的核心问题是：

如何找到既能下降、又不破坏等式约束的牛顿方向？

我们考虑如下的等式约束问题：

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

本节目标

假设 f 是凸函数、问题存在最优解，且相对内点版 Slater 条件成立（例如 f 在 \mathbb{R}^n 上可微且问题可行），则该问题是凸优化问题，并且强对偶成立。

目标： 本节结束后，大家应当能够

- 1 把等式约束问题转换成无约束问题
- 2 写出对偶问题，并说明何时强对偶成立
- 3 从 KKT 条件构造可行/不可行初始点牛顿法

假设与回顾问题

本节统一使用如下假设：

- f 是二阶连续可微凸函数；
- $\mathbf{A} \in \mathbb{R}^{p \times n}$, $\text{rank } \mathbf{A} = p < n$;
- 该优化问题存在最优解；
- 若为保证牛顿步径唯一，默认迭代点附近 $\nabla^2 f(\mathbf{x})$ 在 $\text{Null}(\mathbf{A})$ 上正定。

回顾练习：请写出对该问题的 KKT 条件。

KKT 条件

优化问题

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{Ax} = \mathbf{b} \end{aligned}$$

的最优解满足如下的 KKT 条件:

$$\begin{aligned} \text{(原可行方程)} & & \mathbf{Ax}^* = \mathbf{b}, \\ \text{(驻点方程)} & & \nabla f(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\nu}^* = \mathbf{0}. \end{aligned} \tag{1}$$

例子：一个 KKT 系统怎么解

对于一些特定系统，上一页的 KKT 条件具体化为可求解的线性方程组。考虑最小范数可行点问题：

$$\begin{array}{ll} \text{minimize} & \frac{1}{2} \|\mathbf{x}\|_2^2 \\ \text{subject to} & \mathbf{Ax} = \mathbf{b}, \end{array} \quad \text{rank}(\mathbf{A}) = p.$$

KKT 条件为

$$\begin{bmatrix} \mathbf{I} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \boldsymbol{\nu}^* \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix}.$$

由第一行得 $\mathbf{x}^* = -\mathbf{A}^\top \boldsymbol{\nu}^*$ ，代入第二行：

$$\mathbf{x}^* = \mathbf{A}^\top (\mathbf{AA}^\top)^{-1} \mathbf{b}.$$

对于更一般的情景，我们需要其他工具来实现近似计算。

方法一：转换成无约束问题

方法二：转换成对偶问题

方法三：等式约束的牛顿法

不可行初始点牛顿法

总结

2D 可行集参数化

考虑具体的约束： $\mathbf{x} \in \mathbb{R}^2$, $x_1 + x_2 = 1$ 。我们可以把可行集参数化成

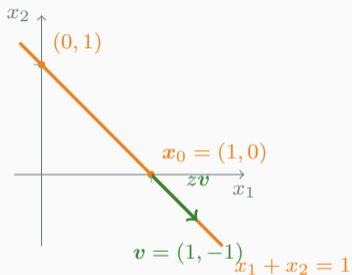
$$\{\mathbf{x} : \mathbf{Ax} = \mathbf{b}\}$$

$$= \{\mathbf{x} : \mathbf{x} = \mathbf{x}_0 + z\mathbf{v} \equiv (1, 0) + z(1, -1)\}$$

故而该优化问题变成了无约束的优化问题

$$\text{minimize } f(\mathbf{x}_0 + z\mathbf{v})$$

因而我们可以使用梯度法、牛顿法等。



一般情景：Null 空间参数化

- 由于矩阵 $\mathbf{A} \in \mathbb{R}^{p \times n}$ 具有 $\text{rank}(\mathbf{A}) = p$ (否则我们可以简化该等式限制), 因此 $\dim(\text{Range}(\mathbf{A}^\top)) = p$ 。因而, 子空间 $\text{Null}(\mathbf{A}) = \text{Range}(\mathbf{A}^\top)^\perp$ 具有维数 $n - p$ 。
- 所以我们可以找到 $n - p$ 个独立的向量 \mathbf{v}_i 使得

$$\text{Null}(\mathbf{A}) = \{\mathbf{F}\mathbf{z} : \mathbf{z} \in \mathbb{R}^{n-p}\}, \quad \mathbf{F} = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_{n-p}].$$

- 总结而言, 我们可以找到 \mathbf{F} 满足

$$\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\} = \{\mathbf{x}_0 + \mathbf{F}\mathbf{z} : \mathbf{z} \in \mathbb{R}^{n-p}\}$$

因此优化问题变成了

$$\text{minimize } f(\mathbf{x}_0 + \mathbf{F}\mathbf{z})$$

链式法则练习

例子： 对前面的 2D 问题 $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{F} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $z \in \mathbb{R}$,

$$\text{minimize } \tilde{f}(z) = f(\mathbf{x}_0 + \mathbf{F}z)$$

请计算 $\tilde{f}'(z)$, $\tilde{f}''(z)$ 。

答案：

$$\tilde{f}'(z) = \mathbf{F}^\top \nabla f(\mathbf{x}_0 + \mathbf{F}z), \quad \tilde{f}''(z) = \mathbf{F}^\top \nabla^2 f(\mathbf{x}_0 + \mathbf{F}z) \mathbf{F}.$$

这里用的是一元链式法则；高维情形完全类似。

无约束形式的梯度与 Hessian

$$\text{minimize } \tilde{f}(z) = f(\mathbf{x}_0 + \mathbf{F}z)$$

通过直接计算可知（具体步骤见课内）：

$$\begin{aligned}\nabla \tilde{f}(z) &= \mathbf{F}^\top \nabla f(\mathbf{x}_0 + \mathbf{F}z) \\ \nabla^2 \tilde{f}(z) &= \mathbf{F}^\top \nabla^2 f(\mathbf{x}_0 + \mathbf{F}z) \mathbf{F}\end{aligned}$$

因此，我们可以直接使用上面的公式来求解下降方向：

$$\text{梯度法} \quad \Delta z = -\nabla \tilde{f}(z)$$

$$\text{牛顿法} \quad \Delta z = -(\nabla^2 \tilde{f}(z))^{-1} \nabla \tilde{f}(z)$$

转换成无约束问题：评价

好处：思路简单

坏处：需要 F ，构建矩阵 F 不一定容易

实际计算中，这个坏处很关键：当约束很多或维数很高时，显式构造 $\text{Null}(\mathbf{A})$ 的一组基可能比直接解 KKT 系统更麻烦。

方法一：转换成无约束问题

方法二：转换成对偶问题

方法三：等式约束的牛顿法

不可行初始点牛顿法

总结

对偶函数的计算

$$\begin{aligned}g(\boldsymbol{\nu}) &= \inf_x f(\boldsymbol{x}) + \boldsymbol{\nu}^\top (\mathbf{A}\boldsymbol{x} - \mathbf{b}) \\&= -\mathbf{b}^\top \boldsymbol{\nu} + \inf_x (f(\boldsymbol{x}) + \boldsymbol{\nu}^\top \mathbf{A}\boldsymbol{x}) \\&= -\mathbf{b}^\top \boldsymbol{\nu} - f^*(-\mathbf{A}^\top \boldsymbol{\nu})\end{aligned}$$

其中 f^* 是 f 的共轭函数:

$$f^*(\mathbf{y}) = \sup_x \{\mathbf{y}^\top \boldsymbol{x} - f(\boldsymbol{x})\}.$$

对偶函数：一个二次例子

若

$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2, \quad f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|_2^2,$$

则等式约束问题

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|\mathbf{x}\|_2^2 \\ &\text{subject to} && \mathbf{Ax} = \mathbf{b} \end{aligned}$$

的对偶函数为

$$g(\boldsymbol{\nu}) = -\mathbf{b}^\top \boldsymbol{\nu} - \frac{1}{2}\|\mathbf{A}^\top \boldsymbol{\nu}\|_2^2.$$

因此，对偶问题是一个关于 $\boldsymbol{\nu}$ 的无约束凹二次最大化问题。

对偶问题：评价

对偶问题是

$$\begin{aligned} & \text{maximize } g(\boldsymbol{\nu}) \\ & = \text{maximize } -\mathbf{b}^\top \boldsymbol{\nu} - f^* (-\mathbf{A}^\top \boldsymbol{\nu}) \end{aligned}$$

好处： 该对偶问题是无约束的优化问题

坏处： 依赖于 f^* 是否好算

方法一：转换成无约束问题

方法二：转换成对偶问题

方法三：等式约束的牛顿法

不可行初始点牛顿法

总结

- 在无约束优化中，我们在 \mathbf{x} 局部使用二次型函数来近似 f ：

$$\hat{f}(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{v} + \frac{1}{2} \mathbf{v}^\top \nabla^2 f(\mathbf{x}) \mathbf{v}$$

其中， $\mathbf{v} = \mathbf{y} - \mathbf{x}$ 。通过最小化 \hat{f} ，我们可得

$$\mathbf{v} = -(\nabla^2 f(\mathbf{x}))^{-1} \nabla f(\mathbf{x})$$

- 或者局部近似梯度

$$\nabla f(\mathbf{y}) \approx \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

通过求解 $\nabla f(\mathbf{y}) = \mathbf{0}$ ，我们依旧可以得到同样的步径 \mathbf{v} 。

(1) 牛顿步径

通过二次型近似的思路，牛顿步径是如下问题中 \mathbf{u} 的最优解

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top \mathbf{u} + \frac{1}{2} \mathbf{u}^\top \nabla^2 f(\mathbf{x}) \mathbf{u} \\ \text{subject to} \quad & \mathbf{A}(\mathbf{x} + \mathbf{u}) = \mathbf{b} \end{aligned}$$

根据该二次型函数的 KKT 条件，最优解满足如下线性方程组：

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}) \\ \mathbf{b} - \mathbf{A}\mathbf{x} \end{bmatrix} \quad (2)$$

在默认正定性假设下，该 KKT 矩阵可解，且 $\Delta \mathbf{x}_{\text{nt}}$ 唯一。

可行点上的牛顿步径

如果当前位置 \boldsymbol{x} 已经满足等式约束 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, 则

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ \mathbf{0} \end{bmatrix} \quad (3)$$

此时第二行给出 $\boldsymbol{A}\Delta \boldsymbol{x}_{\text{nt}} = \mathbf{0}$, 所以牛顿方向沿着可行集的切空间移动。

例子：手算一个可行牛顿步

考虑

$$f(x_1, x_2) = (x_1 - 2)^2 + x_2^2, \quad x_1 + x_2 = 1,$$

从可行点 $\mathbf{x} = (1, 0)^\top$ 开始。此时

$$\nabla f(\mathbf{x}) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \nabla^2 f(\mathbf{x}) = 2\mathbf{I}, \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \end{bmatrix}.$$

可行牛顿步满足

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \omega \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad \Rightarrow \quad \Delta \mathbf{x}_{\text{nt}} = \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}, \quad \omega = 1.$$

所以 $t = 1$ 时, $\mathbf{x} + \Delta \mathbf{x}_{\text{nt}} = (3/2, -1/2)^\top$ 仍满足 $x_1 + x_2 = 1$ 。

通过局部近似梯度

对于公式(1)，我们可以求解在 $(\mathbf{x}, \boldsymbol{\nu})$ 的局部近似解

$$\begin{aligned} \mathbf{A}(\mathbf{x} + \Delta\mathbf{x}_{\text{nt}}) - \mathbf{b} &= \mathbf{0} \\ \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\Delta\mathbf{x}_{\text{nt}} + \mathbf{A}^\top(\boldsymbol{\nu} + \Delta\boldsymbol{\nu}) &= \mathbf{0} \end{aligned}$$

引入符号 $\boldsymbol{\omega} = \boldsymbol{\nu} + \Delta\boldsymbol{\nu}$ 。这里 $\boldsymbol{\omega}$ 只是二次近似问题里的辅助乘子；如果只做可行点牛顿法，我们不需要把它当成下一步真实的 $\boldsymbol{\nu}$ 。该方程等价于

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x}_{\text{nt}} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} -\nabla f(\mathbf{x}) \\ \mathbf{b} - \mathbf{A}\mathbf{x} \end{bmatrix}$$

由此可见，我们得到同样的方向 $\Delta\mathbf{x}_{\text{nt}}$ 。

等式约束的牛顿法

Algorithm 1: 等式约束的牛顿法

Input: 满足等式约束的 $\mathbf{x}^{(0)}$, 阈值 ϵ

```
1 while True do  
2   根据公式(3)计算牛顿步径  $\Delta \mathbf{x}^{(k)}$   
3   计算  $\lambda(\mathbf{x}^{(k)})^2 = (\Delta \mathbf{x}^{(k)})^\top \nabla^2 f(\mathbf{x}^{(k)}) \Delta \mathbf{x}^{(k)}$   
4   停止准则: 如果  $\lambda(\mathbf{x}^{(k)})^2/2 \leq \epsilon$ , 则退出  
5   通过回溯直线搜索选择步长  $t^{(k)} > 0$   
6    $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)} \Delta \mathbf{x}^{(k)}$   
7 end
```

性质：算法能否保持等式约束？

- 如需找到某个 $\boldsymbol{x}^{(0)}$ 满足限制条件，我们仅需找到方程组 $\boldsymbol{Ax}^{(0)} = \boldsymbol{b}$ 的某个解。
- 如果 $\boldsymbol{x}^{(0)}$ 满足 $\boldsymbol{Ax}^{(0)} = \boldsymbol{b}$ ，则后续 $\boldsymbol{x}^{(j)}$ ($j \geq 0$) 永远满足等式约束，无论步长 $t^{(j)}$ 怎么选。
- 假使 $\boldsymbol{x}^{(0)}$ 不可行，若用公式(2)计算方向并取 $t^{(0)} = 1$ ，则 $\boldsymbol{x}^{(1)}$ 满足等式约束；但这一步未必留在 f 的定义域内。

性质：算法能否保证函数值下降？

对于任意的定义域内的 \mathbf{x} ，我们可以计算在牛顿方向，函数值变化如下：

$$\begin{aligned}\frac{d}{dt}f(\mathbf{x} + t\Delta\mathbf{x}_{nt})\Big|_{t=0} &= \nabla f(\mathbf{x})^\top \Delta\mathbf{x}_{nt} \\ &= -\Delta\mathbf{x}_{nt}^\top (\nabla^2 f(\mathbf{x})\Delta\mathbf{x}_{nt} + \mathbf{A}^\top \boldsymbol{\omega}) \\ &= -\Delta\mathbf{x}_{nt}^\top \nabla^2 f(\mathbf{x})\Delta\mathbf{x}_{nt} + \boldsymbol{\omega}^\top (\mathbf{A}\mathbf{x} - \mathbf{b})\end{aligned}$$

此处，我们使用了公式(2)来简化。

当 \mathbf{x} 满足等式限制 $\mathbf{A}\mathbf{x} = \mathbf{b}$ 时，

$$\frac{d}{dt}f(\mathbf{x} + t\Delta\mathbf{x}_{nt})\Big|_{t=0} = -\Delta\mathbf{x}_{nt}^\top \nabla^2 f(\mathbf{x})\Delta\mathbf{x}_{nt} \leq 0.$$

在 $\text{Null}(\mathbf{A})$ 上的正定性假设下，若 $\Delta\mathbf{x}_{nt} \neq \mathbf{0}$ ，则该方向是严格下降方向。

方法 1 和方法 3 的关联：方向

对于思路 1：我们需要求解如下的问题

$$\text{minimize } \tilde{f}(z) = f(\mathbf{x}_0 + \mathbf{F}z)$$

其中牛顿步径为

$$\begin{aligned}\Delta z &= -(\nabla^2 \tilde{f}(z))^{-1} \nabla \tilde{f}(z) \\ &= -\left(\mathbf{F}^\top \nabla^2 f(\mathbf{x}_0 + \mathbf{F}z) \mathbf{F}\right)^{-1} \mathbf{F}^\top \nabla f(\mathbf{x}_0 + \mathbf{F}z)\end{aligned}$$

因此，对于 $\mathbf{x} = \mathbf{x}_0 + \mathbf{F}z$ ， \mathbf{x} 方向的牛顿步径为

$$\Delta \mathbf{x} = \mathbf{F} \Delta z = -\mathbf{F} \left(\mathbf{F}^\top \nabla^2 f(\mathbf{x}) \mathbf{F}\right)^{-1} \mathbf{F}^\top \nabla f(\mathbf{x})$$

方法 1 和方法 3 的关联：结论

当 \boldsymbol{x} 满足等式限制 $\boldsymbol{Ax} = \boldsymbol{b}$ 时，方法 1 给出的方向

$$\Delta \boldsymbol{x} = -\boldsymbol{F}(\boldsymbol{F}^\top \nabla^2 f(\boldsymbol{x}) \boldsymbol{F})^{-1} \boldsymbol{F}^\top \nabla f(\boldsymbol{x})$$

也是方程(2)的解，即

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^\top \\ \boldsymbol{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x} \\ \boldsymbol{\nu} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}) \\ \mathbf{0} \end{bmatrix}.$$

因此方法 1 和方法 3 的牛顿法其实是等价的，但方法 3 不需要提前构造 \boldsymbol{F} 。

局部收敛性

若最优解附近的 KKT 矩阵非奇异， $\nabla^2 f$ 在可行切空间上正定，并且 $\nabla^2 f$ 在邻域内 Lipschitz 连续，则当初始点足够接近最优解且最终接受 $t = 1$ 时，该等式约束牛顿法具有局部二次收敛性。

证明思路见附录。课堂中更重要的是：两种推导给出同一个 $\Delta \boldsymbol{x}$ 。

方法一：转换成无约束问题

方法二：转换成对偶问题

方法三：等式约束的牛顿法

不可行初始点牛顿法

总结

不可行初始点：动机

问题来源： 等式约束的牛顿法本身很高效，一个小问题是：如果函数本身的定义域不是 \mathbb{R}^n ，有时可能求解某个 \mathbf{x} 来满足 $\mathbf{Ax} = \mathbf{b}$ 未必很容易；而使用牛顿法中 $t = 1$ 来使得等式约束满足也未必可行。

问题： 在不强制等式约束的情况下，能否依旧构造牛顿法？

考虑一般的 KKT 条件(1)，我们可以求解在 $(\mathbf{x}, \boldsymbol{\nu})$ 的局部近似解

$$\begin{aligned} \mathbf{A}(\mathbf{x} + \Delta\mathbf{x}_{\text{nt}}) - \mathbf{b} &= \mathbf{0} \\ \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\Delta\mathbf{x}_{\text{nt}} + \mathbf{A}^\top(\boldsymbol{\nu} + \Delta\boldsymbol{\nu}_{\text{nt}}) &= \mathbf{0} \end{aligned} \tag{4}$$

我们把 $(\mathbf{x}, \boldsymbol{\nu})$ 看成是变量，而不是单纯仅更新 \mathbf{x} 。

不可行初始点：方向方程

从(4)整理可得，方向 $(\Delta \mathbf{x}_{\text{nt}}, \Delta \boldsymbol{\nu}_{\text{nt}})$ 满足

$$\begin{bmatrix} \nabla^2 f(\mathbf{x}) & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_{\text{nt}} \\ \Delta \boldsymbol{\nu}_{\text{nt}} \end{bmatrix} = - \begin{bmatrix} \nabla f(\mathbf{x}) + \mathbf{A}^\top \boldsymbol{\nu} \\ \mathbf{A}\mathbf{x} - \mathbf{b} \end{bmatrix}.$$

和可行点牛顿法相比，右端多了当前的原始残差 $\mathbf{A}\mathbf{x} - \mathbf{b}$ 和对偶残差 $\nabla f(\mathbf{x}) + \mathbf{A}^\top \boldsymbol{\nu}$ 。

课堂练习：不可行残差怎么变

若 $\mathbf{x}^{(k)}$ 不可行，且方向 $\Delta \mathbf{x}_{\text{nt}}^{(k)}$ 由不可行初始点牛顿方程得到，请证明

$$\mathbf{A}(\mathbf{x}^{(k)} + t\Delta \mathbf{x}_{\text{nt}}^{(k)}) - \mathbf{b} = (1 - t)(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}).$$

请思考：

- 当 $t = 1$ 时，等式约束残差会发生什么？
- 为什么实际算法中仍然需要回溯直线搜索？

答案提示： 由方向方程第二行可得 $\mathbf{A}\Delta \mathbf{x}_{\text{nt}}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}$ 。

由不可行初始点牛顿方程的第二行可得

$$\mathbf{A}\Delta\mathbf{x}_{\text{nt}}^{(k)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}.$$

因此对任意步长 t ,

$$\begin{aligned}\mathbf{A}(\mathbf{x}^{(k)} + t\Delta\mathbf{x}_{\text{nt}}^{(k)}) - \mathbf{b} &= \mathbf{A}\mathbf{x}^{(k)} - \mathbf{b} + t\mathbf{A}\Delta\mathbf{x}_{\text{nt}}^{(k)} \\ &= \mathbf{A}\mathbf{x}^{(k)} - \mathbf{b} + t(\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}) \\ &= (1 - t)(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{b}).\end{aligned}$$

所以若 $t = 1$, 原可行残差变为 $\mathbf{0}$ 。但实际中仍需回溯直线搜索, 因为 $\mathbf{x}^{(k)} + \Delta\mathbf{x}_{\text{nt}}^{(k)}$ 可能不在 f 的定义域内。

不可行初始点：待解决的问题

从方程(4)得到方向后，一个自然的迭代骨架是：

1. 由当前 $(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)})$ 计算方向 $(\Delta \mathbf{x}_{\text{nt}}^{(k)}, \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)})$;
2. 选择步长 $t^{(k)} > 0$;
3. 更新 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)} \Delta \mathbf{x}_{\text{nt}}^{(k)}$, $\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + t^{(k)} \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)}$ 。

新问题： 因为涉及到了新变量 $\boldsymbol{\nu}$,

- 回溯直线搜索的标准是什么？
- 终止条件是什么？

不可行初始点：残差函数

可以考虑如下残差。先分别定义

$$r_{\text{dual}}(\mathbf{x}, \boldsymbol{\nu}) = \left\| \nabla f(\mathbf{x}) + \mathbf{A}^{\top} \boldsymbol{\nu} \right\|_2,$$
$$r_{\text{pri}}(\mathbf{x}) = \left\| \mathbf{A}\mathbf{x} - \mathbf{b} \right\|_2,$$

再令

$$r(\mathbf{x}, \boldsymbol{\nu}) = \sqrt{r_{\text{dual}}(\mathbf{x}, \boldsymbol{\nu})^2 + r_{\text{pri}}(\mathbf{x})^2},$$
$$R(\mathbf{x}, \boldsymbol{\nu}) = r(\mathbf{x}, \boldsymbol{\nu})^2.$$

其中 r_{dual} 衡量驻点条件是否满足， r_{pri} 衡量等式约束是否满足。

不可行初始点：残差下降性

则对于点 $(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)})$ ，通过方程(4)来计算牛顿方向 $(\Delta \mathbf{x}_{\text{nt}}^{(k)}, \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)})$ ，可以验证

$$\left. \frac{d}{dt} R(\mathbf{x}^{(k)} + t\Delta \mathbf{x}_{\text{nt}}^{(k)}, \boldsymbol{\nu}^{(k)} + t\Delta \boldsymbol{\nu}_{\text{nt}}^{(k)}) \right|_{t=0} = -2R(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)}) \leq 0$$

【证明见该 PPT 的附录。】

当 $r(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)}) > 0$ 时，也有 $r'(0) = -r(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)})$ 。因此，我们可以使用 r 做回溯直线搜索。

不可行初始点牛顿法

Algorithm 2: 不可行初始点牛顿法

Input: 定义域内的 $\mathbf{x}^{(0)}, \boldsymbol{\nu}^{(0)}$; 阈值 ϵ, η ; 回溯参数 α, β

```
1 while True do
2     解线性方程(4), 得到方向  $\Delta \mathbf{x}_{\text{nt}}^{(k)}, \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)}$ 
3     若  $r_{\text{dual}}(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)}) \leq \epsilon$  且  $r_{\text{pri}}(\mathbf{x}^{(k)}) \leq \eta$ , 则退出
4     令  $t = 1$ 
5     不断令  $t = \beta t$ , 直到定义域条件和残差下降条件同时满足
6     令  $t^{(k)} = t$ 
7      $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)} \Delta \mathbf{x}_{\text{nt}}^{(k)}, \boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + t^{(k)} \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)}$ 
8 end
```

其中残差下降条件为 $r(\mathbf{x}^{(k)} + t \Delta \mathbf{x}_{\text{nt}}^{(k)}, \boldsymbol{\nu}^{(k)} + t \Delta \boldsymbol{\nu}_{\text{nt}}^{(k)}) \leq (1 - \alpha t) r(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)})$ 。

性质：不可行残差如何变化？

假使 $\mathbf{x}^{(0)}$ 不满足 $\mathbf{Ax}^{(0)} = \mathbf{b}$ ，并且我们一直取 $t^{(k)} \leq 1$ ，则

$$\mathbf{Ax}^{(k+1)} - \mathbf{b} = \mathbf{Ax}^{(k)} + t^{(k)} \mathbf{A}\Delta\mathbf{x}_{\text{nt}}^{(k)} - \mathbf{b} = (1 - t^{(k)}) (\mathbf{Ax}^{(k)} - \mathbf{b})$$

因此，残差 $\|\mathbf{Ax}^{(k+1)} - \mathbf{b}\| = \left(\prod_{j=0}^k (1 - t^{(j)}) \right) \|\mathbf{Ax}^{(0)} - \mathbf{b}\|$ 按该乘积因子衰减；若步长有正下界，则几何衰减。

课堂快速检验

考虑二次目标

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{H} \mathbf{x} + \mathbf{c}^\top \mathbf{x}, \quad \mathbf{A} \mathbf{x} = \mathbf{b},$$

其中 \mathbf{H} 在 $\text{Null}(\mathbf{A})$ 上正定。

请独立完成：

1. 当 $\mathbf{x}^{(k)}$ 可行时，写出求 $\Delta \mathbf{x}^{(k)}$ 的 KKT 线性方程组；
2. 用一行式子说明为什么 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t \Delta \mathbf{x}^{(k)}$ 仍然可行；
3. 若 $\mathbf{x}^{(k)}$ 不可行，写出原-对偶残差 $r(\mathbf{x}^{(k)}, \boldsymbol{\nu}^{(k)})$ 。

答案提示： 分别对应公式(3)、 $\mathbf{A} \Delta \mathbf{x}^{(k)} = 0$ 、以及不可行初始点牛顿法中的残差。

方法一：转换成无约束问题

方法二：转换成对偶问题

方法三：等式约束的牛顿法

不可行初始点牛顿法

总结

主要需要掌握的知识:

- 掌握三种不同的思路来解等式约束的优化问题
- 重点掌握公式(2), (3), (4), 即牛顿步径的定义
- 能够复述写出等式约束的牛顿法
- 能够复述写出不可行初始点的牛顿法

阅读作业 & 参考资料:

- 课本第 10 章

附录：方法 1 和方法 3 为什么等价

记 $H = \nabla^2 f(\mathbf{x})$, $\mathbf{g} = \nabla f(\mathbf{x})$, 且

$$\Delta \mathbf{x} = -\mathbf{F}(\mathbf{F}^\top \mathbf{H} \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{g}.$$

由于 $\mathbf{A} \mathbf{F} = \mathbf{0}$, 所以 $\mathbf{A} \Delta \mathbf{x} = \mathbf{0}$, KKT 系统的第二行成立。

第一行只需证明

$$\mathbf{H} \Delta \mathbf{x} + \mathbf{g} \in \text{Range}(\mathbf{A}^\top).$$

由于 $\text{Null}(\mathbf{A}) = \text{Range}(\mathbf{A}^\top)^\perp$, 等价于

$$\mathbf{F}^\top (\mathbf{H} \Delta \mathbf{x} + \mathbf{g}) = \mathbf{0}.$$

代入 $\Delta \mathbf{x}$ 即可得到该等式。因此存在 $\boldsymbol{\nu}$ 使得 KKT 系统第一行成立。

附录：残差下降性（设置）

考虑求解 $\mathbf{g}(\mathbf{y}) = \mathbf{0}$ ，其中

$$\mathbf{g}(\mathbf{y}) = \begin{bmatrix} g_1(\mathbf{y}) \\ \vdots \\ g_m(\mathbf{y}) \end{bmatrix} \in \mathbb{R}^m.$$

假设 $\nabla \mathbf{g}(\mathbf{y})$ 可逆。通过局部线性化，牛顿方向满足

$$\mathbf{g}(\mathbf{y}) + \nabla \mathbf{g}(\mathbf{y}) \Delta \mathbf{y} = \mathbf{0}.$$

也就是说，

$$\nabla \mathbf{g}(\mathbf{y}) \Delta \mathbf{y} = -\mathbf{g}(\mathbf{y}).$$

附录：残差下降性（验证）

令 $\phi(t) = \|\mathbf{g}(\mathbf{y} + t\Delta\mathbf{y})\|^2$ ，则

$$\begin{aligned}\phi'(0) &= 2\mathbf{g}(\mathbf{y})^\top \nabla \mathbf{g}(\mathbf{y}) \Delta\mathbf{y} \\ &= -2\mathbf{g}(\mathbf{y})^\top \mathbf{g}(\mathbf{y}) \\ &= -2\|\mathbf{g}(\mathbf{y})\|^2 \leq 0.\end{aligned}$$

因此，该方向是残差函数 $\|\mathbf{g}(\mathbf{y})\|^2$ 的下降方向。

附录：应用到 KKT 残差

之前 KKT 条件可以变成该形式的特例： $m = n + p$,

$$\mathbf{y} = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\nu} \end{bmatrix} \in \mathbb{R}^{n+p},$$

$$\begin{bmatrix} g_1(\mathbf{y}) \\ \vdots \\ g_p(\mathbf{y}) \end{bmatrix} = \mathbf{A}\mathbf{x} - \mathbf{b},$$

$$\begin{bmatrix} g_{p+1}(\mathbf{y}) \\ \vdots \\ g_{n+p}(\mathbf{y}) \end{bmatrix} = \nabla f(\mathbf{x}) + \mathbf{A}^\top \boldsymbol{\nu}.$$

并且，残差 $R(\mathbf{x}, \boldsymbol{\nu}) = \|\mathbf{g}(\mathbf{y})\|_2^2$ 。